WHAT IS CLAIMED IS:

Sub5A1

1. A method for the cloning of intact, diversity-selected genes from within gene cassettes, said method comprising the steps of:

(a) identifying repeat DNA sequences which flank gene cassettes;

(b) hybridizing oligonucleotides to said repeated sequences which flank said gene cassettes and amplifying said sequences to provide DNA fragments which contain genes from within the cassettes.

(c) ligating said DNA fragments into a vector; and

(d) transforming said vector into an appropriate strain.

2. The method of claim 1 wherein said diversity-selected genes are selected from the group consisting of:

cell surface antigens such as polysaccharide antigens or polypeptide antigens or secreted molecules; adhesins such as fimbrial proteins, pilus proteins or outer membrane proteins; transporters of small molecules, especially those with narrow specificity; toxins, hemolysins, hemagglutins, kinases and signaling molecules;

detoxifying enzymes such as drug resistance determinants; catabolic enzymes specific for compounds episodically available, excluding those required for central metabolic pathways such as the tricarboxylic acid cycle; enzymes for biosysnthesis of rare sugars, excluding those required in all cells, such as ribose, deoxyribose, and sugars of the cell wall, especially of those sugars that form part of the pericellular envelope.

3. The method of claim 2 wherein said diversity-selected genes comprise restriction endonuclease genes.

4. The method of claim 2 wherein said diversity-selected genes comprise methyltransferase genes.

5. The method of claim 1 wherein said oligonucleotides contain recognition sites which permit directional cloning.

6. The method of claim 5 wherein the DNA fragments are ligated into said vector in an orientation that enables expression.

7. A method for identifying the presence of gene cassette arrays from within a target DNA preparation, said method comprising the steps of:

(a) hybridizing at least one oligonucleotide which hybridizes to one or more of SEQ ID NO:5 through SEQ ID NO:78 to a DNA preparation; and

(b) detecting the presence of a stable DNA-DNA hybrid.

8. The method of claim 7 wherein said detection comprises determining the presence of stable DNA-DNA hybrid by Southern blot or dot blot.

9. The method of claim 7 wherein said detection comprises employing at least two oligonucleotides and hybridizing said oligonucleotides to said DNA preparation, and detecting their ability to support DNA polymerization at the 3' end of the stable DNA-DNA hybrid.

10. The method of claim 7 wherein said oligonucleotides comprise SEQ ID NO:79 through SEQ ID NO:91.

11. The method of claim 7 wherein said oligonucleotides hybridize to one or more of DNA SEQ ID NO:5 through SEQ ID NO:78 or portions thereof.

12. The method of claim 7 wherein the DNA source comprises an individual strain.

13. The method of claim 7 wherein the DNA source comprises a group or pool of strains.

14. The method of claim 7 wherein the DNA source comprises environmental DNA.

15. A composition consisting of isolated DNA primers comprising SEQ ID NO:79 through SEQ ID NO:91 or portions thereof.

16. A composition consisting of DNA primers which hybridize to one or more of DNA SEQ ID NO:5 through SEQ ID NO:78 or portions thereof.

17. A method for identifying gene cassette arrays from a predetermined DNA sequence, said method comprising the steps of:

(a) screening the said predetermined DNA sequence for TAACWA;

(b) screening the said predetermined DNA sequence for CGTTRR;

(c) screening for DNA segments wherein the 5' T of step A is less than about 200 base pairs form the 3' R of step B; and

(d) determining whether the DNA sequence of step C is repeated in the predetermined DNA sequence.